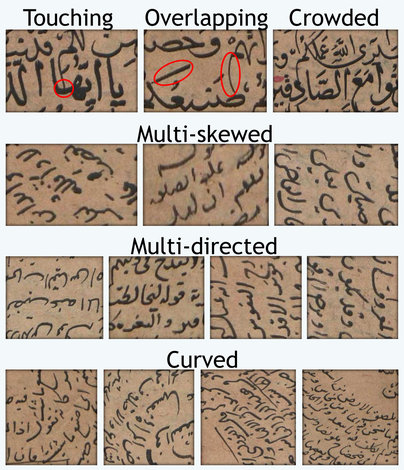
Text Line Segmentation for Challenging Handwritten Document Images Using Fully Convolutional Network

Ben-Gurion University of the Negev

Berat Barakat Ahmad Droby Majeed Kassis Jihad El-Sana

Introduction

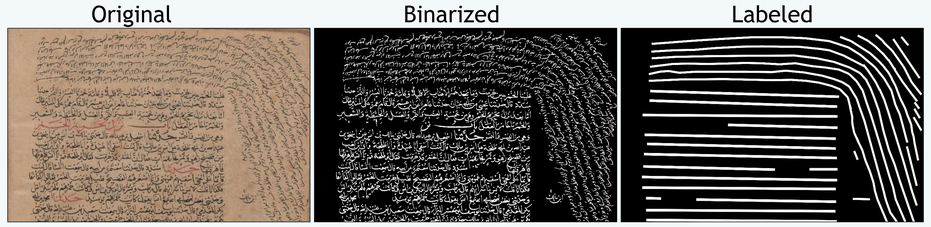


Challenging handwritten documents contain various writing styles with inconsistent font types and font sizes through multi-skewed, multi-directed and curved text lines.

This paper

1. Provides a dataset of challenging documents.
2. Describes text line segmentation of this dataset using Fully Convolutional Network (FCN).
3. Proposes a new metric that is sensitive to both over and under segmentation of lines.

Data and Method



Dataset

The challenging dataset contains 30 pages from two different manuscripts. It is written in Arabic language and contains 2732 text lines. We applied 6-fold cross validation. Each fold was split into train, validation and test sets.

Pre-processing

1. Binarize and invert document images
2. Manually label line masks on binarized document images
3. Generate 50.000 random patches of size 320×320 for training
4. Generate 6.000 random patches of size 320×320 for validation

FCN architecture

We used the FCN proposed for semantic segmentation [1]. FCN inputs the original images and their pixel level annotations for learning the hypothesis function that can predict whether a pixel belongs to a text line label or not.
The crucial question is how to annotate the text lines.

Line mask labeling connects the characters in the same line.

Advantages:

1. Is applicable to all the alphabets in contrast to baseline labeling
2. Is not cumbersome for crowded documents in contrast to bounding polygon

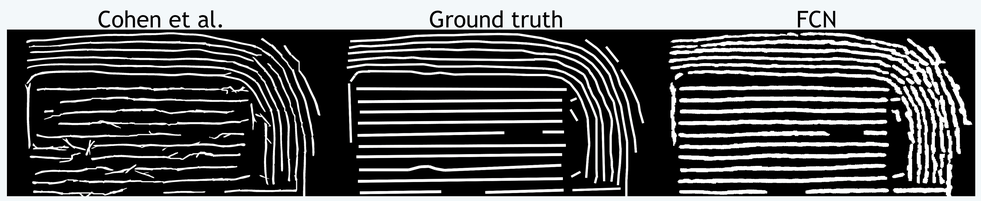
Disadvantages:

1. Disregards diacritics and touching components between lines.

Results

Post-processing

Occasionally predicted line masks were disconnected. First, orientation of each connected component was computed. Then a directional morphological operation was applied.



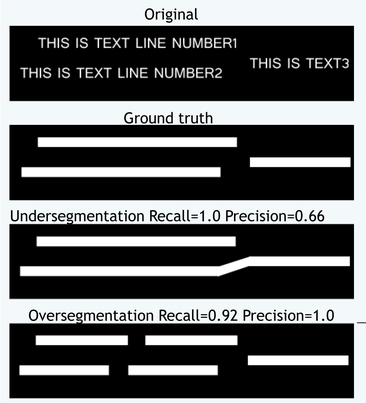
Testing

During the testing, a sliding window of size 320×320 was used for prediction, but only the inner window of size 100×100 was considered.

Error analysis

Most errors of FCN method occur at curved areas whereas most errors of method of Cohen et al. occur at the main text areas. The former was a result of small number of training patches with curved lines. The latter was a result of biased average character height.

Evaluation



New metric

Correct extraction: All connected components of a line with the same label
Under segment: Some connected components of a line with the same label
Over segment: All connected components of a line with multiple labels

	Proposed	Cohen et al.
Recall	**0.82**	0.74
Precision	**0.78**	0.60
F-measure	**0.80**	0.66

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.